Facultat de Ciències

Incidence of stroke in Lleida (2010-2014)

Final degree project

Author:
Ramon Boix
NIU: 1334215

Supervisor:
Joan Valls

September 4, 2016

Abstract

In this work, we present a statistical and epidemiological analysis of the stroke incidence, being currently the second leading cause of global death and cause of disability in adults worldwide level. More particularly, we have analysed its recent trend up to 2014, for all the socio-health care region in Lleida. This research project reveals a reasonably accurate stroke current situation in Lleida, being this level of interest not only epidemiological but also in terms of public health and proper adequacy of resources, economic and healthcare, to face the diagnosis, treatment and prevention of this disease.
Regarding statistical methodology, we provide an efficient modeling frame to analyse incidence data for a given disease, by means of age-period models. All the computation work have been performed with R statistical package.

Keywords: *Stroke, Incidence rate, Poisson models, Epidemiology, Annual Percent Change, Lleida.*

Resum

En aquest treball, presentem un anàlisi estadístic i epidemiològic de la incidència de l'íctus que actualment és la segona causa de mort i de discapacitats en població adulta a nivell mundial. Més concretament, hem analitzat la seva tendència fins al 2014 per a la regió socio-sanitària de Lleida. Aquest projecte d'investigació mostra la situació actual de l'íctus, de manera raonablement acurada a Lleida. L'interès d'aquest no només és a nivell epidemiològic, sinó també en termes de salut pública i correcte gestió dels recursos econòmics i sanitàris, per a fer front a les diagnosis, pel tractament i la prevenció de la malaltia.
Pel que fa a la metodologia estadística, treballem en un marc eficient per analitzar dades d'incidència, donada una malaltia. S'han emprat models d'edat-període amb aquesta finalitat. Tots els càlculs s'han realitzat emprant el paquet estadístic R.

Paraules clau: *Íctus, Taxa d'Incidència, Models de Poisson, Epidemiologia, Percentatge de canvi anual, Lleida.*

Resumen

En este trabajo, se presenta un análisis estadístico i epidemiológico de la incidencia del ictus, siendo esta la segunda causa de muerte i de discapacidades en población adulta a nivel mundial. Más concretamente, se han analizado las tendencias hasta el 2014 para la región sociosanitaria de Lérida. Este proyecto de investigación muestra la situación actual del ictus, de manera razonablente cuidadosa en Lérida. El interés de este no solo es a nivel epidemiológico, sino también en términos de salud pública y una gestión de los recursos económicos i sanitarios adecuados, para hacer frente al diagnóstico, para el tratamiento y la prevención de la enfermedad.
Respecto a la metodología estadística, se ha trabajo en un marco eficiente para analizar datos de incidencia, dada una enfermedad. Se han utilizado modelos edad-período con este objectivo. Todos los cálculos se han realizado con el paquete estadístico R.

Palabras clave: *Ictus, Tasa de Incidencia, Modelos de Poisson, Epidemiología, Porcentage de cambio anual, Lérida.*

Contents

1 Preface

When most of the pople think about statistics, they tend to visualise it as a useless tool. In fact, people gets disconcerted when I introduce myself as a statistician. What it surprises me, is that they are not able to see how important could be this field in our society. Statistics, in a nutshell, is the skill of understanding, processing, exploring and eventually be able to summarise and understand the insights of any given data. These days, I believe we live in the data age, where either the public and the private sector are generating an enormous amounts of data. In this sense, I belive that statisticians will be able to make a great contribution.
I always thought, as an analogy, that statisticians are like journalists. I will explain, journalists are instructed on communication skills, to be objectives, ... but they can apply all those things in such different areas. Statisticians are taught on programming skills, quantitative vision, applied mathematics, ... but they can as well apply all those things in many areas. In my case, I decided to start my professional career in biomedical research. The work presented in this document, is the culmination of several months of my internship at the Institut de Recerca Biomèdica de Lleida (IRBLleida)[1] . During this time, I was working at the Unitat de Bioestadística i Epidemiologia (UBE)[2] and I have been involved in several biomedical research analysis with different multidisciplinar groups.
Regarding the analysis be exposed on this document, it was headed by Dr. Francesc Purroy as principal investigator and head of the Clinical Neuroscience group[3] . Ana Belén Vena and Joan Valls were my mentors in the medical and the methodological parts, correspondingly. The study, aims to describe the incidence of the Stroke disease in Lleida. If I would be forced to sum up the whole work in two words I may choose: *Stroke* and *Epidemiology*. *Stroke*, of course, is the disease which is the object of the study and is broadly known for most of the population. Nevertheless, there are few studies that quantify the incidence or mortality in Spain. Some recent studies have tried to collect data for the incidence of stroke in different countries. In that sense, underdeveloped countries have less control on this matter, making much difficult to obtain a worlwide stroke incidence estimation.
On the other hand, *epidemiology*,[4] is the study and analysis of the patterns, causes/effects or changes of disease conditions in defined populations. They silently work in issues of public health, aiming to identify risk factors, health policies strategies and tracking their effects along time. For that matter, they help to correctly design studies, statistical analysis and interpretation of the results. For all these reasons I feel specially close to epidemiologists and it was such a pleasure to analise a database that could contribute somehow in the scientifical comunity.
I want to specially thank for all the statistical knowledge that I have been thaught and for making my work so much easier with the guidance of my final degree project tutor, Joan Valls [7]. Thank you for the time you invest in me, for the patience and all the support you have given to me. Finally, I would like to thank all the people I worked with during my stage at IRBLleida and have given me the opportunity to develop my skills in data analysis.

(a) Composition of the "Unitat de bioestadística i epidemiologia" at IRBLleida

(b) Head of the Neuroscience research group at IRBLleida

[1] `http://www.irblleida.org`
[2] `https://www.facebook.com/BiostatsLleida/`
[3] `https://clinicalneurosciencesIleida.wordpress.com/`
[4] *Epidemiology*, literally meaning "the study of what is upon the people", is derived from Greek: *epi* meaning "upon", *demos* meaning "people" and *logos* meaning "study".

2 Introduction

The work presented in this project is within the framework of the descriptive analysis of epidemiological data on Stroke. More particularly, the analysis in centered on the assessment on the incidence of Stroke in Lleida from 2010 to 2014 period and its trend. The incidence will be studies globally and taking into account age groups[5]. The data will be studied taking into account the sex and specific stroke subtypes.
The stroke is a cardiovascular accident. It happens when blood flow to a portion the brain is impaired due to a blood clot or broken blood vessel. Much like a heart attack, the lack of oxigen-rich blood can lead to tissue death. When brain cells begin to die as a result, symptoms occur in the parts of the body that those brain cells control. These symptoms can include sudden weakness, paralysis,numbness of face or limbs, and difficulty to think, move or even breath. It always have been a lot of controversy on the understanding of the disease. The first stroke recorded in history was recognised by *Hippocrates*, known to be the *father of medicine*, more than 2.400 years ago. He called the condition apoplexy, which is in Greek terms that stands for "struck down by violence". Back in the seventeenth century, a doctor named Jacob Wepfer discovered that some of the apoplexy cases were massive bleeding into the brain, while in other the arteries were blocked.
Nowadays, it is known by the medical community that stroke can be separated into two groups. Stroke can be ischemic type -basically due to a loss or reduction of blood supply to the brain or haemorrhagic type -due to haemorrhagic rupture of a cerebral vessel-. Regarding its type, Ischemic stroke have an estimated prevalence of 85% of the registered cases. Haemorrhagic stroke, despite a lower prevalence presented -around 15% of cases of stroke- is associated with mortality and a worse prognosis.

In terms of epidemiology, stroke is now, according to the World Health Organization[6], the second overall cause of death in the world and the leading cause of disability in adults. Epidemiological studies conducted to date have shown that the incidence increases in developed countries due course to an increase in life expectancy, being considered a problem of public health and being highly prevalent in the elder population (greater than 80 years), where it became the principal health problem. The health and economic burden of disease is also very important, estimated by the Centers of Disease Control and prevention (CDC)[7] with a total annual costs of 34 billion dollars.

In Catalonia, although some studies have been performed [8], there is no systematic estimation of the incidence of stroke in the population or its recent trend. To the best of our knowledge, the results of this study would be able to assess -the impact of the most recent health policies[8]. An example of another international health policies related to stroke is the *ACT F.A.S.T* [2] campaign. Launched by the British health authorities to quickly detect the sympthoms of the stroke and call emergency services.

Figure 1: Campaign launched by the catalan health authorities in order to sensibilise the population about the stroke disease.

The structure of the document could be summarised in two main blocks: First, all the methodological framework used for the analysis will be displayed, followed by the results of the research project analysis and its accurate interpretation.

[5] e.g. {<40}, {40-54},{55-70}, {71-80} and {>80}
[6] `http://www.who.int/mediacentre/factsheets/fs310/en/`
[7] `http://www.cdc.gov/stroke/facts.htm`
[8] In 2006, began the campaign called "Codi ictus" designed by the catalan health authorities. The model implemented was an emergency code to quickly identify the simpthoms of the stroke.

3 Statistical methods

3.1 Incidence Rate

Incidence rate [3] (the so-called *"crude rate"*) from a certain disease, like Stroke, have been used as populational indicators in epidemiologics in order to analise its trend and to evaluate the impact of new treatments or other health programs implementations. Usually, the incidence of a certain disease can be aggregated by age groups. Denote with i the different age groups, $i = 1, \dots, I$ and t the period of time in which this rates have been observated. For our porpouses, we will consider $I = 5$ that corresponds to the following five age groups: $\{¡40\}, \dots, \{¿80\}$ and t is considered to be an integer that refers to a certain year.
Let us denote c_it the number of cases for the age group i and period t. Let's assume that c_{it} is a positive integer. On the other hand, let n_{it} to be the number of people at risk on the period t for the same age group i. Then, the incidence rate is defined as:

$$M_{it} = \frac{c_{it}}{n_{it}}$$

Assume that λ_{it} is the probability to develope a certain disease for the age group i during period t. Thus, is usually called the *risk* of a certain disease. In terms of statistics, c_it is considered to be a random variable with expectancy $E[c_{it}] = \lambda_{it} n_{it} = \theta_{it}$. More specifically, the number of cases is considered to be characterised by a Poisson distribution:

$$c_{it} \sim \mathcal{P}(\theta_{it} = \lambda_{it} n_{it})$$

Then, M_it is an estimator of λ_{it} and can be denoted as $M_{it} = \hat{\lambda_{it}}$. In practice, incidence rates are given by 100,000 people, such as $r_{it} = \frac{c_{it}}{n_{it}} 100.000$ for a given period t and the age group i. It could be interesting to have an estimator for all the population, r_i, that can be written as:

$$r_t = \frac{\sum_{i=1}^{I} c_{it}}{\sum_{i=1}^{I} n_{it}}$$

As stated before, number of cases are assumed to be distributed as a Poisson random variable $c_{it} \sim \mathcal{P}(\lambda_{it} n_{it})$, being submitted to a certain variance. In order to evaluate that variance in our estimations, it is a must to present the rates with their corresponding confidence intervals. Poisson distribution is known to have the same mean and variance as:

$$\mathrm{E}(c_{it}) = \mathrm{Var}(c_{it})$$

Then, we can obtain the variance of the incidence risk M_{it} as:

$$\mathrm{Var}\left(\frac{c_{it}}{n_{it}}\right) = \frac{\mathrm{Var}(c_{it})}{n_{it}^2} = \frac{\lambda_{it}}{n_{it}}$$

And the variance of the age adjusted rates can be obtained as:

$$\mathrm{Var}(M_t^{adj}) = \mathrm{Var}(w_t M_{it}) = w_t^2 \mathrm{Var}(M_{it}) = w_t^2 \frac{\lambda_i t}{n_i t}$$

Given that λ_{it} is always an unkown parameter, as stated before can be estimated it with $\hat{\lambda_{it}} = \frac{c_{it}}{n_{it}}$. At this point, risk estimator M_t is assumed to be assimptotically normally distributed. The reason is because it is a sum of independent random variables. Then the distribution of M_t is:

$$M_{it} \sim \mathrm{N}\Big(\hat{\lambda_t}, \mathrm{Var}(\hat{M_{it}})\Big)$$

If we scale the normal variable, we can obtain a confidence interval for $\rho_t = \lambda_t 100.000$ with $1 - \alpha$ confidence level as:

$$\rho_t \in \left[r_t - z_{1-\frac{\alpha}{2}} 10^5 \sqrt{\mathrm{Var}(M_{it})} \quad , \quad r_t + z_{1-\frac{\alpha}{2}} 10^5 \sqrt{\mathrm{Var}(M_{it})} \right]$$

Using the same arguments, the age-specific confidence intervals can be computed subsetting the sample to a certain age group i. It is worth comment, that main assumption on this procedure is the normal approximation of the rates. Thus, in practice requires that sample size is *big enough*. As there is no convention to know when this criteria is met and sould always be judged by expertise criteria. Furthermore, the Poisson distribution assumption of the number of cases c_{it} should be checked. In practise, the most common deviation from the Poisson distribution of the cases could be caused due to overdispersion ($\hat{\sigma^2} > \hat{\mu}$) on the sample.

3.2 Age adjusted rate

Sometimes investigators are interested in comparing rates in different years, regions or times. In that cases, the age of the population could have an insight effect on the comparative of the rates. In other words, the age structures between populations can be the cause of the observed differences. That's the main reason why incidence (or mortality) rates should be standardised when your main interest is comparing it with others.
The most common method is the *direct method* to adjust the incidence rates. Let's assume we have a standard population1, more specifically, the proportional weights w_i for each age group i. Thus, $\sum_{i=1}^{I} w_i = 1$ and the age adjusted rates (ASR's) [3] are:

$$r_t^{adj} = \sum_{i=1}^{I} w_i \frac{c_{it}}{n_{it}} 100.000 = \sum_{i=1}^{I} w_i r_i t$$

Using the arguments exposed in Section 3.1, if we substitute $\mathrm{Var}(M_{it})$ for $\mathrm{Var}(M_{it}^{adj})$ we obtain the confidence interval for the age adjusted rates.

3.3 Age-period model

As we seen before c_{it} and n_{it} are the number of cases and the population at risk for a given age group i and a certain period t. The most common model used to assess the trend is the age-period model [4] and can be written as:

$$\log\left(\mathrm{E}\Big(\frac{c_{it}}{n_{it}}\Big)\right) = \alpha_i + \beta_i t$$

The model assume that the risk to develop certain disease is log-linear related with time. For the linear predictors (α_i) and their corresponding slopes (β_i), each age group have its own specific effect. This model, can be integrated in the framework of the generalised linear models [5], with the logarithmic link and the Exponential family distribution equals to Poisson.
In order to evaluate the trend of the rates for a given disease, the most usuall measure is the Annual Percent Change (APC) that quantifies the intensity of the annual increase/decrease of the rates. Let $\rho_t = r_t 100.000$, then the formal definition of the APC is:

$$\mathrm{APC}_i = \frac{\lambda_{it} - \lambda_{i(t-1)}}{\lambda_{i(t-1)}}$$

With the assumption that the APC remain constant through time and it is different for each age group.
From the model in Section 3.3, can be deduced that $\lambda_{ij} = e^{\alpha_i + \beta_i t}$ from which allows us to rewrite the formula of the APC as follows:

$$\mathrm{APC}_i = \frac{e^{\alpha_i+\beta_i t} - e^{\alpha_i+\beta_i(t-1)}}{e^{\alpha_i+\beta_i(t-1)}} = e^{\beta_i} - 1$$

Then, using the properties of the maximum likelihood estimators, an estimator of the annual percent change can be obtained as $\mathrm{APC}_i = e^{\hat{\beta}_i} - 1$, where β_i is the maximum likelihood estimation for the parameter on the model in Section 3.3. The variance of β_i can be obtained through the model standard estimation as $\mathrm{Var}(\beta_i) = \sigma^2_{\beta_i}$. Using the properties of the maximum likelihood estimators, we can assume that β_i is assimptotically normal distributed. Therefore, scaling the distribution of the parameter β_i we obtain the confidence interval for the APC such as:

$$\mathrm{APC}_i \in \left[e^{\hat{\beta}_i - z_{\frac{\alpha}{2}}\hat{\sigma}^2_{\beta_i}} - 1 \quad , \quad e^{\hat{\beta}_i + z_{\frac{\alpha}{2}}\hat{\sigma}^2_{\beta_i}} - 1 \right]$$

Finally, in some cases is interesting to adjust the same slope for each age group. Thus, can be interpreted as there are no differences in the trend for all age groups. In such cases, the most common model used is the *"age-drift"* [4] that can be written as follows:

$$\log\left(\mathrm{E}\Big(\frac{c_{it}}{n_{it}}\Big)\right) = \alpha + \beta t$$

Using the same arguments used in the *"age-period"* model, APC and their corresponding confidence intervals can be computed.

4 Research project

The research project that will be exposed was developed by the Clinical Neuroscience grup at IRBLleida. The project was named "Incidence of Stroke in Lleida from 2010 to 2014 period". Furthermore, this is a test driver to apply for a medicine research grant that allows us to replicate using data from the whole catalan territory.. First, we will explain about the design of the study, followed by the methodology and finally, we will show the results of the analysis.

4.1 Design of the study

Observational, population based and cross-sectional study, including patients diagnosed with a Stroke, as primary or secundary diagnosis in hospitalary discharges registered in Lleida from 2010 to 2014. The inclusion codes (ICD-10), that will quantify the cases, are those from 430 to 432 (haemorrhagic stroke), from 433 to 435 (ischemic stroke). The data was provided by the informatics of the Hospital Arnau de Vilanova. With regards to the population of Lleida, the data have been obtained from IDESCAT.

4.2 Statistical methods

The methodology used in this project is in line with another national studies with similar characteristics [1].First of all, the crude incidence rates will be calculated and then adjusted by age (using the standard population from World Health Organisation 2001 1). The estimations of the rates will be performed in number of cases per 100.000 people-year, computing the rates globally, separately by sex, age-groups and Stroke sub-type. Log-linear models based on the Poisson distribution will be used to estimate the variability and their corresponding confidence intervals. Age-Period Poisson models with log link will be used to compute the APC (Annual Percent Change) and their corresponding confidence intervals, including statisticall significance of the change. All the analysis are performed with R statistical package and taking the statistical significance level at 5% ($\alpha = 0.05$), therefore, the estimations for the confidence intervals at 95% confidence.

4.3 Objectives

- To determine the incidence of Stroke in Lleida from 2010 to 2014 and their respective time trend. Globally, stratifying by Stroke sub-types (Ischemic or Haemorrhagic), Sex and Age-groups.
- To compare the global incidence rate obtained in Lleida with different regions around the world.
- To create a scientific-technical report with the epidmiological and statistical information that will allow to evaluate the situation of the Stroke in Lleida and be useful for the implementation of Public health measures, as well as to set up the right policies of the echonomics and sanitary resources, improve the diagnosis, treatments and the prevention actions of the Stroke.

4.4 Statistical analysis

4.4.1 Graphical representations

Figures 6, 7 and 8 revealed the incidence in stroke in Lleida (from 2010 to 2014 period) in Male, Female and All population both global and considering age groups, taking into account the Global stroke or stroke subtypes (Ischemic or Haemorrhagic). These figures can be found in the Supplementary materials (Section 6.1) and we will use its content later on this document.

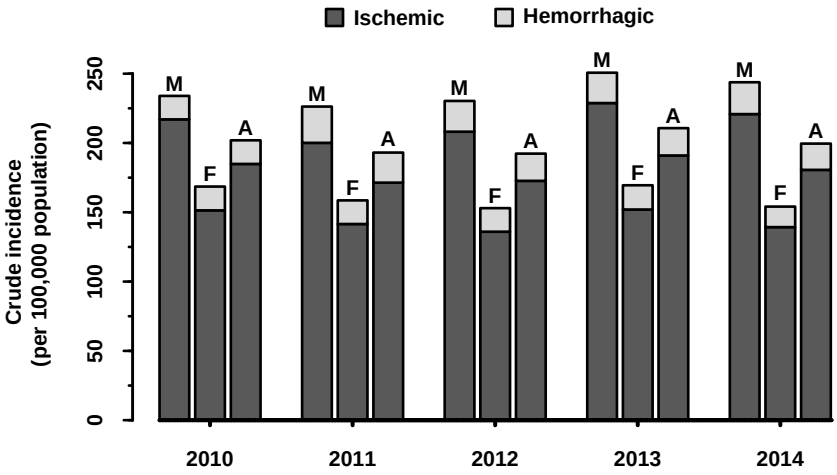

Figure 2: *Evolution of the Stroke incidence in Lleida (2010 to 2014).* Trend in the Stroke incidence shown, using the Crude Rate (per 100.000) for Male, Female and All population. Decomposition of the Crude Rate corresponding into Ischemic and Haemorrhagic stroke subtypes also displayed.

First of all, we will be focused on the crude rate of the global population[9]. In that sense, Figure 2 depicts the global incidence rate in Lleida from 2010 to 2014 period. Visual inspection showed no important changes in the trend of the incidence rates from 2010 to 2014 (considering: men, women and all population). If we consider stroke subtypes, it can be seen that Ischemic stroke represent most of the cases. Again, no significant trends can be appreciated in the incidence rates from 2010 to 2014 (considering: men, women and all population).
It can be seen that the incidence of stroke its not the same when taking into account the sex. Indeed, the incidence is higher in male population.

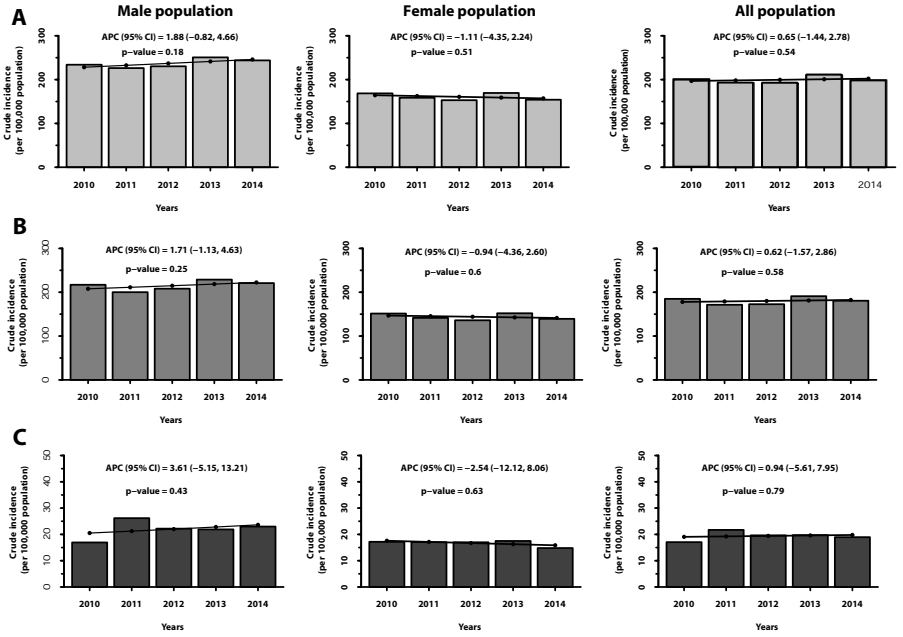

Figure 3: *Assessment of the trend in Stroke incidence in Lleida (2010-2014).* Crude incidence per 100,000 from 2010 to 2014 period shown for sex and stroke subtype. Male, Female and All population displayed in the first, second and third columns; respectively. Global stroke and their corresponding subtypes (Ischemic and Haemorrhagic) shown in rows A, B and C; respectively. Annual Percent Change (APC) with 95% confidence interval displayed. P-values assessing the statistical signification of the trend obtained from Poisson generalised linear models with logarithmic link. Prediction obtained from models shown in dots and lines.

[9]e.g. When considering all the population together and do not take into account the age groups.

Now, we are interested in analise the time trend of the incidence in stroke. Our purpose is to use statistics methods [10] to decide if the incidence rate is increasing/decreasing or is stable. For that issue, age-period models were used. More precisely, on Figure 3 contains the trend amongst the sex and the stroke subtypes. It is worth mention, even though no statistical significant trends were found, the trends were increasing and decreasing for male and female population, correspondingly. The highest APC (with 95% confidence interval) was found in Haemorrhagic stroke subtype and male population: APC = 3.61% (-5.15%, 13.21%). The results indicated that the trend on the incidence rate in stroke was stable (even when considering sex and stroke subtypes).

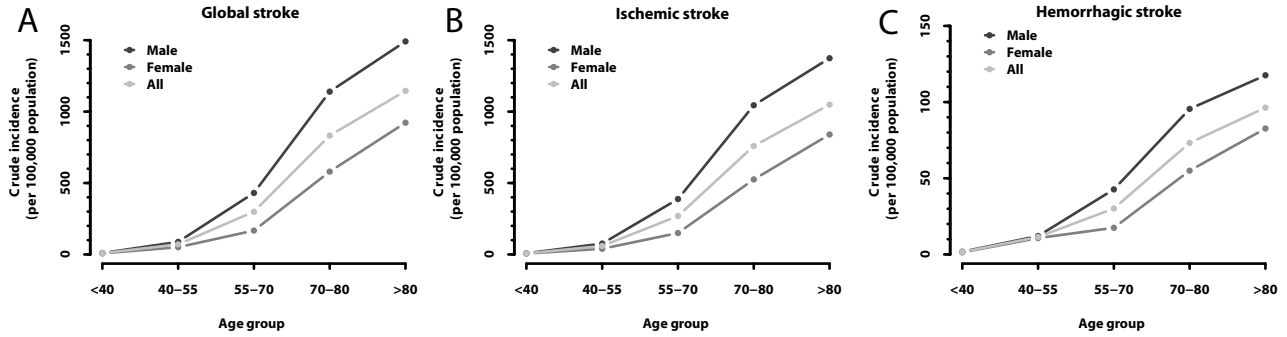

Figure 4: *Assessment of the trend in Stroke incidence in Lleida (2010-2014) for age groups.* Trend in the Stroke incidence shown, using the Crude Rate (per 100,000) for Male, Female and All population for each age group threshold (<40, 40-54, 55-70, 71-80, >80). Global stroke and their corresponding subtypes (Ischemic and Haemorrhagic) displayed from left to right; in A, B and C, correspondingly.

It is known that stroke incidence is not the same accross age groups. According to medical criteriun, it was decided to use the following age groups for the porpouses of this analysis: {<40, 40-55, 56-70, 71-80, >80}. Figure 4 depicts the average[11] crude rate in stroke for sex and stroke subtypes. The incidence accross the age groups was found much higher for Ischemic than for Haemorrhagic stroke subtype. It can be seen a clear association between the age and the incidence rate. More precisely, the higher the age group is, the higher the crude become. This association was in line with the scientific literature. Again, men showed higher incidence rate than women for each age group and stroke subtype. No differences can be observer with regards to the behaviour of the trends in different stroke subtypes.

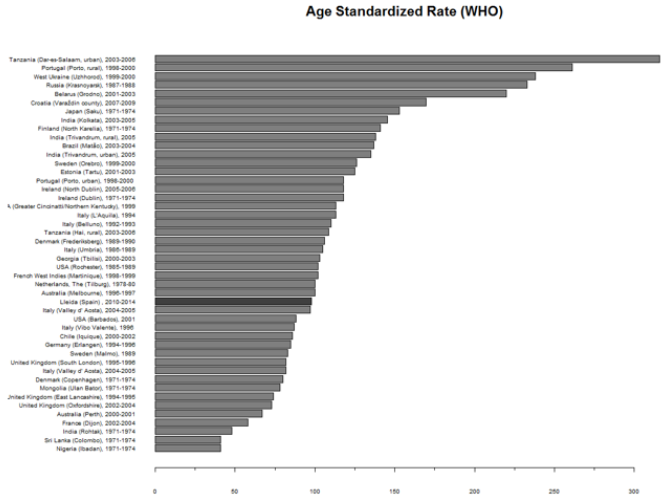

Figure 5: *Global stroke incidence.* Age Standardized Rate per 100,000 (World Health Organization 2001 standard population) shown for several worldwide regions and periods. Estimation for the incidence in Lleida with darker grey shown

.

[10] Jointpoint regression models have been carried out to challenge if there is a change on the trend. No statistical results were obtained

[11] Since no statistical significant trends were found, it was decided to represent the average of the crude rate between 2010 to 2014 for each age group.

It was explained, in Section 3.2 that ASR's (age standardized rates) were used to compare incidence rates coming from different populations. To achieve that purpouse, *direct method* was used. It is important to remark ASR's are used to rank rates from diferent populations and can not be interpreted for its number[12]. Figure 5 showed the ASR in stroke for several regions around the world [6]. In general terms, the ASR of stroke in Lleida is whithin the expected range, being this result lower than the middle position of the figure.

4.4.2 Results of the analysis

Graphical representations of the analysis were shown in Section 4.4.1, nevertheless an overall view of the results will be displayed in this Section. The structure is composed by three blocks (All stroke subtypes, Ischemic and Haemorrhagic stroke subtype) followed by a general review of the analysis:

- When considering all the stroke subtypes, the observed cases were: n = 888, n= 854, n = 852, n = 929 and n = 874 for 2010 to 2014 period, correspondingly. The incidence of stroke is found higher in male (n = 525, n = 510, n = 519, n = 561 and n = 541) than female (n = 363, n = 344, n = 333, n = 368 and n = 333) population. The estimation of the global trend over time (with 95% confidence interval) were: APC = 1.88% (-0.82%, 4.66%), APC = -1.11% (-4.35%, 2.24%) and APC = 0.65% (-1.44%, 2.78%) for Male, Female and All population, respectively. None of them was found statistically significance, so there is no evidence to reject the stability assumption for the incidence of global stroke. If we consider 5 age groups (<40, 40-55, 56-70, 71-80, >80), the incidence rate of the global stroke increases with the age (as can be seen in Figure 4). The trend in the 71-80 age group for Female population was found APC = -9.47% (-14.90%, -3.73%) and showed a clear decrease of approximately 10% yearly (p-value = 0.002) in the incidence rate. Figure **??** depicts the trend along age groups in global stroke, revealing a decrease of the APC when increase the age in Males and an increasing of the APC when increasing the age in Females.

- When considering Ischemic stroke subtype, the observed cases were: n = 813, n= 758, n = 765, n = 842 and n = 791 for 2010 to 2014 period, correspondingly. The incidence of stroke is found higher in male (n = 487, n = 451, n = 469, n = 512 and n = 490) than female (n = 326, n = 307, n = 296, n = 330 and n = 301) population. The estimation of the global trend over time (with 95% confidence interval) were: APC = 1.71% (-1.13%, 4.63%), APC = -0.94% (-4.36%, 2.60%) and APC = 0.62% (-1.57%, 2.86%) for Male, Female and All population, respectively. None of them was found statistically significance, so there is no evidence to reject the stability assumption for the Ischemic stroke incidence. If we consider 5 age groups (<40, 40-55, 56-70, 71-80, >80), the incidence rate of the global stroke increases with the age (as can be seen in Figure 4). The trend in the 71-80 age group for Female population was found APC = -8.72% (-14.46%, -2.64%) and showed a clear decrease of approximately 9% yearly (p-value = 0.006) in the incidence rate. **??** depicts the trend along age groups in global stroke, revealing a decrease of the APC when increase the age in Male and an increasing of the APC when increasing the age in Female population.

- When considering Haemorrhagic stroke subtype, the observed cases were: n = 75, n= 96, n = 87, n = 87 and n = 83 for 2010 to 2014 period, correspondingly. The incidence of stroke is found higher in male (n = 38, n = 59, n = 50, n = 49 and n = 51) than female (n = 37, n = 37, n = 37, n = 38 and n = 32) population. It is worth mention that Haemorrhagic stroke, is the minority subtype and only represent the 10% of the global stroke cases (as can be seen in Figure 1). The estimation of the global trend over time (with 95% confidence interval) were: APC = 3.61% (-5.15%, 13.21%), APC = -2.54% (-12.12%, 8.06%) and APC = 0.94% (-5.62%, 7.95%) for Male, Female and All population, respectively. None of them was found statistically significance, so there is no evidence to reject the stability assumption for the Ischemic stroke incidence. If we consider 5 age groups (<40, 40-55, 56-70, 71-80, >80), the incidence rate of the global stroke increases with the age (as can be seen in Figure 4). The trend in the age group was not found statistically significant, so there is no evidence to reject the stability assumption for the incidence rate. Figure **??** depicts the trend along age groups in global stroke, revealing no clear patterns nor in Male or Female population.

[12] An ASR = 95 should not be interpreted as 95 incident cases per 100.000 people. Its only value is to rank populations from different regions, sex or periods.

- In general terms, it can be seen a higher incidence of stroke in Male than Female population (even when considering different stroke subtypes). The incidence rate is found to increase when the age increases (even when considering stroke subtypes). The statistical power is deficient when analysing Haemorrhagic stroke subtype (due to lower incidence) and the estimations are not as accurate as the estimations in global stroke or Ischemic stroke subtype. The incidence of stroke (globally and Ischemic subtype) on female population between 71-80 years old are found to be decreasing approximately 10% yearly, being this result statistically significant. With regards to the global trend of stroke is found to be stable in Male, Female and All population (even when considering different subtypes). The estimated trend in age groups decreases when the age increases in Male population and increases when the age increases in Female population (when considering Global and Ischemic stroke subtype). Figure 5 compare the estimated global Age standardised rate (ASR = 97.83 [77.85, 117.81]) per 100,000 population found in Lleida from 2010 to 2014 with other regions worldwide, showing a plausible result.

5 Conclusions

In the introduction of this work we exposed the origins of this project, which began when we worked in a biomedical research project at IRBLleida under the supervision of Joan Valls. That project was designed to study the incidence of the stroke in Lleida from 2010 to 2014 period. Formal definition and the history of the stroke disease were provided on the begining of this work. We have been able to define and set which are the main objectives of the epidemiology. In a nutshell, *epidemiology* is closed related with medicine but is more focused on quantify which is the impactt of certain diseases into the whole population. In order to achieve that goal, *epidemiologists* need to have a strong statistical background.

We have introduced the most common statistics to study the incidence/death for a given disease, setting the theoretical framework that we used in the statistical analysis. That is, the incidence rate, age standardised rates and the *Age-period model* to evaluate their time trends. We defined their respective confidence intervals to properly study the variance of our estimators. Our reference in the methodological approach is known to be the most used amongst the scientifical community [3].

We have defined the characteristics of the study, the methodological approach and its objectives. The study is observational, retrospective and population-based. Regarding to the statistical framework used, we formally defined in Section 6.1. On the other hand there are three main objectives of the research project: -To assess the incidence of the stroke in Lleida from 2010 to 2014 (stratifying by sex, age groups and stroke subtype), -To assess their trend over time and -To compare the incidence found with other regions of the world.

Some graphical representations were displayed to correctly interpret the results of the statistical analysis in Section 4.4.1. Furthermore, we summarised all the important information in Figures 6, 7 and 8 and discused the overall results in 4.4.2. The incidence of Stroke in Lleida from 2010 to 2014 period was characterised by higher incidence in male population, an increasement of the incidence with the age of the patients. When considering stroke subtypes, Ischemic stroke was found the most common (more than 90% of the cases) and the rest corresponds to Haemorrhagic stroke. Regarding to the trend of the incidence, there was no statistical significance evidence to reject the stability of the incidence from 2010 to 2014[13]. Despite the fact that there were no statistical significant trends, one can not omit, in terms of descriptive statistics, that the APC was increasing and decreasing; for male and female population, respectively. Thus, indicates in alignment of the medical community that no epidemiological outbreaks happenned from 2010 to 2014 in the incidence of stroke in Lleida. No differences found in the trends with regards to different stroke subtypes.

Finally, we have been able to compare the global age standardised rate estimated in Lleida with different regions around the world, using the most recent world estimation [6]. The estimated global ASR in Lleida was slightly below the average of the other regions, indicating that there are no big differences between the analised regions due to demographics or another risk factors. All in all, the Neuroscience research grup at IRBLleida and us considered that the analysis results succesfully reflects the objectives of the work. Also, this research project served as the pilot test to be applied to the whole catalan territory, conditioned to the medical funding grants available.

With regards to the statistical methodology in this paper, we tried other models that have not been presented due to lack of adjustment. For the *Age-period* and the *Age-draft* models, identity link and Gaussian distribution had been performed. With regards to the trend, the adjustment of the APC obtained by the *Age-period* models assume that there is no change[14] in the whole period. For this purpose, *jointpoint regression* models have been performed with no statistical significant results. Perhaps, with a bigger period on our data, some changes on the trends could have been detected.

[13] Except for Female population with Global or Ischemic stroke subtype within the 70-80 age group.

[14] e.g. from 2010 to 2012 increasing incidence rate and from 2013 to 2015 decreasing rates

6 Annex

6.1 Supplementary materials

6.1.1 Global stroke

All stroke subtypes

Male	Age groups					Global		
Year	<40 cases (Crude rate)	40-55 cases (Crude rate)	55-70 cases (Crude rate)	70-80 cases (Crude rate)	>80 cases (Crude rate)	cases	Crude rate (95% CI)	Adjusted rate (95% CI)
2010	5 (4.32)	42 (81.57)	129 (409.26)	178 (1179.2)	171 (1632.93)	525	233.96 (214.37,254.86)	129.75 (118.18,142.83)
2011	18 (15.66)	52 (98.94)	129 (400.45)	157 (1059.45)	154 (1416.87)	510	226.28 (207.06,246.79)	130.08 (118.17,143.5)
2012	9 (7.95)	43 (80.59)	140 (420.71)	154 (1066.78)	173 (1556.18)	519	230.31 (210.92,251.01)	126.55 (115.19,139.39)
2013	10 (9.05)	44 (81.81)	166 (487.23)	173 (1223.39)	168 (1482.27)	561	250.69 (230.38,272.32)	140.63 (128.51,154.23)
2014	14 (13.03)	53 (98.08)	152 (435.7)	164 (1170.43)	158 (1368.91)	541	243.81 (223.7,265.25)	135.5 (123.63,148.86)
			time trend				time trend	
APC (95% CI)	11.53 (-7.40, 34.68)	1.82 (-7.02, 11.53)	3.27 (-1.95, 8.78)	1.24 (-3.52, 6.24)	-3.05 (-7.63, 1.76)		1.88 (-0.82, 4.66)	
p-value	0.25	0.7	0.22	0.62	0.21		0.18	

Female	Age groups					Global		
Year	<40 cases (Crude rate)	40-55 cases (Crude rate)	55-70 cases (Crude rate)	70-80 cases (Crude rate)	>80 cases (Crude rate)	cases	Crude rate (95% CI)	Adjusted rate (95% CI)
2010	10 (9.63)	23 (51.01)	59 (185.3)	121 (660.16)	150 (921.04)	363	168.55 (151.65,186.81)	72.96 (64.44,83.1)
2011	7 (6.75)	22 (47.66)	53 (164.35)	124 (685.99)	138 (821.18)	344	158.58 (142.27,176.26)	67.68 (59.62,77.31)
2012	8 (7.77)	17 (36.27)	47 (142.52)	104 (587.01)	157 (915.83)	333	152.97 (136.98,170.32)	61.58 (53.88,70.85)
2013	10 (9.87)	29 (61.42)	57 (169.07)	87 (503.44)	185 (1051.91)	368	169.48 (152.6,187.71)	68.73 (60.45,78.59)
2014	4 (4.03)	27 (56.72)	60 (175.41)	79 (462.96)	163 (902.2)	333	154.09 (137.98,171.56)	61.26 (53.66,70.46)
			time trend				time trend	
APC (95% CI)	-9.99 (-28.35, 12.52)	-1.11 (-4.35, 2.24)	-0.78 (-8.73, 7.85)	-9.47 (-14.90, -3.73)	2.1 (-2.81, 7.27)		-1.11 (-4.35, 2.24)	
p-value	0.36	0.51	0.85	0.002	0.41		0.51	

All	Age groups					Global		
Year	<40 cases (Crude rate)	40-55 cases (Crude rate)	55-70 cases (Crude rate)	70-80 cases (Crude rate)	>80 cases (Crude rate)	cases	Crude rate (95% CI)	Adjusted rate (95% CI)
2010	15 (6.83)	65 (67.31)	188 (296.72)	299 (894.57)	321 (1199.64)	888	201.92 (188.86,215.65)	99.42 (92.29,107.32)
2011	25 (11.44)	74 (74.96)	182 (282.33)	281 (854.23)	292 (1055.14)	854	193.08 (180.34,206.47)	97.61 (90.41,105.56)
2012	17 (7.87)	60 (59.87)	187 (282.24)	258 (802.41)	330 (1167.73)	852	192.31 (179.61,205.67)	92.48 (85.65,100.04)
2013	20 (9.44)	73 (72.28)	223 (328.99)	260 (827.45)	353 (1220.57)	929	210.7 (197.37,224.69)	102.74 (95.47,110.75)
2014	18 (8.71)	80 (78.71)	212 (306.84)	243 (781.95)	321 (1084.13)	874	199.54 (186.53,213.22)	96.64 (89.63,104.39)
			time trend				time trend	
APC (95% CI)	2.01 (-11.50, 17.78)	2.93 (-4.42, 10.86)	2.28 (-2.14, 5.90)	-3.00 (-6.61, 0.74)	-0.57 (-3.95, 2.92)		0.65 (-1,44, 2.78)	
p-value	0.78	0.45	0.32	0.12	0.74		0.54	

Figure 6: *Incidence of stroke in Lleida (2010-2014).* Number of stroke cases, Crude Rate (CR) per 100,000 in Lleida from 2010 to 2014 displayed for age groups for male female and both sex population. Global estimation of the cases, CR (95% confidence interval) or ASR (95% confidence interval) for all years also shown for male, female and both sex population. WHO (World Health Organization 2001) standard population used to compute the Age standardized Rate. Annual Percent Change (APC) with 95% confidence interval shown for male , female and both sex population. Poisson models with logarithmic link used to assess the trend and their corresponding statistical significance.

6.1.2 Ischemic stroke

Ischemic stroke

Male	Age groups					Global		
Year	<40 cases (Crude rate)	40-55 cases (Crude rate)	55-70 cases (Crude rate)	70-80 cases (Crude rate)	>80 cases (Crude rate)	cases	Crude rate (95% CI)	Adjusted rate (95% CI)
2010	5 (4.32)	39 (75.75)	116 (368.02)	166 (1099.7)	161 (1537.43)	487	217.03 (198.18,237.19)	119.76 (108.67,132.35)
2011	16 (13.92)	38 (72.3)	115 (356.99)	144 (971.73)	138 (1269.67)	451	200.1 (182.06,219.45)	114.1 (103.02,126.69)
2012	7 (6.19)	36 (67.47)	128 (384.65)	141 (976.72)	157 (1412.25)	469	208.13 (189.71,227.84)	114.25 (103.47,126.51)
2013	7 (6.34)	41 (76.23)	151 (443.21)	158 (1117.32)	155 (1367.57)	512	228.8 (209.41,249.5)	127.49 (116.03,140.44)
2014	11 (10.24)	48 (88.83)	135 (386.97)	148 (1056.24)	148 (1282.27)	490	220.83 (201.71,241.28)	121.99 (110.76,134.71)
			time trend				time trend	
APC (95% CI)	5.32 (-14.23,29.48)	4.11 (-5.587,14.85)	3.21 (-2.28,9.02)	0.51 (-4.43,5.70)	-2.92 (-7.69,2.10)		1.71 (-1.13, 4.63)	
p-value	0.62	0.42	0.26	0.84	0.25		0.24	
Female	**Age groups**					**Global**		
Year	<40 cases (Crude rate)	40-55 cases (Crude rate)	55-70 cases (Crude rate)	70-80 cases (Crude rate)	>80 cases (Crude rate)	cases	Crude rate (95% CI)	Adjusted rate (95% CI)
2010	8 (7.7)	17 (37.71)	54 (169.60)	108 (589.23)	139 (853.49)	326	151.37 (135.38,168.72)	64.14 (56.26,73.65)
2011	6 (5.79)	18 (39)	49 (151.94)	108 (597.48)	126 (749.78)	307	141.53 (126.14,158.28)	59.9 (52.37,69.02)
2012	7 (6.8)	13 (27.74)	41 (124.33)	98 (553.14)	137 (799.16)	296	135.98 (120.92,152.38)	54.53 (47.33,63.32)
2013	7 (6.91)	24 (50.83)	50 (148.31)	82 (474.51)	167 (949.57)	330	151.98 (136.02,169.29)	60.4 (52.73,69.66)
2014	4 (4.03)	21 (44.11)	53 (154.94)	70 (410.22)	153 (846.85)	301	139.28 (123.99,155.94)	54.05 (46.96,62.74)
			time trend				time trend	
APC (95% CI)	-9.43 (-29.61, 15.86)	6.46 (-7.83,23.14)	-2.06 (-10.33,6.98)	-8.72 (-14.46,-2.64)	2.28 (-2.87,7.71)		-0.94 (-4.36, 2.60)	
p-value	0.43	0.4	0.64	0.006	0.39		0.6	
All	**Age groups**					**Global**		
Year	<40 cases (Crude rate)	40-55 cases (Crude rate)	55-70 cases (Crude rate)	70-80 cases (Crude rate)	>80 cases (Crude rate)	cases	Crude rate (95% CI)	Adjusted rate (95% CI)
2010	13 (5.92)	56 (57.99)	170 (268.31)	274 (819.77)	300 (1121.16)	813	184.87 (172.38,198.03)	90.15 (83.39,97.66)
2011	22 (10.07)	56 (56.73)	164 (254.41)	252 (766.07)	264 (953.96)	758	171.37 (159.39,184.02)	85.78 (79.08,93.23)
2012	14 (6.48)	49 (48.89)	169 (255.08)	239 (743.32)	294 (1040.34)	765	172.67 (160.65,185.36)	82.88 (76.43,90.06)
2013	14 (6.61)	65 (64.36)	201 (296.53)	240 (763.80)	322 (1113.38)	842	190.97 (178.28,204.31)	92.21 (85.37,99.78)
2014	15 (7.26)	60 (67.88)	188 (272.1)	218 (701.51)	301 (1016.58)	791	180.59 (168.23,193.63)	86.45 (79.85,93.8)
			time trend				time trend	
APC (95% CI)	-1.03 (-15.50, 15.85)	4.83 (-3.32, 13.70)	1.88 (-2.74, 6.73)	-3.11 (-6.88, 0.81)	-0.44 (-3.96, 3.22)		0.62 (-1.57, 2.86)	
p-value	0.9	0.25	0.43	0.12	0.81		0.58	

Figure 7: *Incidence of Ischemic stroke in Lleida (2010-2014).* Number of stroke cases, Crude Rate (CR) per 100,000 in Lleida from 2010 to 2014 displayed for age groups for male female and both sex population. Global estimation of the cases, CR (95% confidence interval) or ASR (95% confidence interval) for all years also shown for male, female and both sex population. WHO (World Health Organization 2001) standard population used to compute the Age standardized Rate. Annual Percent Change (APC) with 95% confidence interval shown for male , female and both sex population. Poisson models with logarithmic link used to assess the trend and their corresponding statistical significance.

6.1.3 Haemorrhagic stroke

Hemorrhagic stroke

Male	Age groups					Global		
Year	<40 cases (Crude rate)	40-55 cases (Crude rate)	55-70 cases (Crude rate)	70-80 cases (Crude rate)	>80 cases (Crude rate)	cases	Crude rate (95% CI)	Adjusted rate (95% CI)
2010	0 (0)	3 (5.83)	13 (41.24)	12 (79.5)	10 (95.49)	38	16.93 (11.98,23.24)	10 (6.92,14.92)
2011	2 (1.74)	14 (26.64)	14 (43.46)	13 (87.73)	16 (147.21)	59	26.18 (19.93,33.77)	15.98 (11.86,21.87)
2012	2 (1.77)	7 (13.12)	12 (36.06)	13 (90.05)	16 (143.92)	50	22.19 (16.47,29.25)	12.3 (8.94,17.4)
2013	3 (2.72)	3 (5.58)	15 (44.03)	15 (106.07)	13 (114.7)	49	21.9 (16.2,28.95)	13.14 (9.45,18.6)
2014	3 (2.79)	5 (9.25)	17 (48.73)	16 (114.19)	10 (86.64)	51	22.98 (17.11,30.22)	13.51 (9.91,18.85)
			time trend				time trend	
APC (95% CI)	47.88 (-6.26,149.52)	-11.52 (-31.22, 13.18)	3.84 (-11.93, 22.61)	9.64 (-7.21, 29.77)	-4.57 (-19.75, 13.40)		3.61 (-5.151,13.21)	
p-value	0.11	0.33	0.65	0.28	0.59		0.43	

Female	Age groups					Global		
Year	<40 cases (Crude rate)	40-55 cases (Crude rate)	55-70 cases (Crude rate)	70-80 cases (Crude rate)	>80 cases (Crude rate)	cases	Crude rate (95% CI)	Adjusted rate (95% CI)
2010	2 (1.93)	6 (13.31)	5 (15.7)	13 (70.93)	11 (67.54)	37	17.18 (12.1,23.68)	8.82 (5.84,13.81)
2011	1 (0.96)	4 (8.67)	4 (12.40)	16 (88.52)	12 (71.41)	37	17.06 (12.01,23.51)	7.77 (5.14,12.36)
2012	1 (0.97)	4 (8.53)	6 (18.19)	6 (33.87)	20 (116.67)	37	17.00 (11.97,23.43)	7.05 (4.56,11.5)
2013	3 (2.96)	5 (10.59)	7 (20.76)	5 (28.93)	18 (102.35)	38	17.50 (12.38,24.02)	8.33 (5.47,13.16)
2014	0 (0)	6 (12.6)	7 (20.46)	9 (52.74)	10 (55.35)	32	14.81 (10.13,20.9)	7.21 (4.68,11.73)
			time trend				time trend	
APC (95% CI)	-12.49 (-50.30, 48.76)	0.7 (-23.85, 33.34)	10.81 (-14.40, 44.47)	-16.34 (-32.01, 2.23)	0.29 (-14.94, 18.33)		-2.54 (-12.12,8.06)	
p-value	0.62	0.96	0.44	0.09	0.97		0.63	

All	Age groups					Global		
Year	<40 cases (Crude rate)	40-55 cases (Crude rate)	55-70 cases (Crude rate)	70-80 cases (Crude rate)	>80 cases (Crude rate)	cases	Crude rate (95% CI)	Adjusted rate (95% CI)
2010	2 (0.91)	9 (9.32)	18 (28.41)	25 (74.8)	21 (78.48)	75	17.05 (13.41,21.38)	9.28 (7.1,12.34)
2011	3 (1.37)	18 (18.23)	18 (27.92)	29 (88.16)	28 (101.18)	96	21.70 (17.58,26.50)	11.83 (9.32,15.21)
2012	3 (1.39)	11 (10.98)	18 (27.17)	19 (59.09)	36 (127.39)	87	19.64 (15.73,24.22)	9.6 (7.46,12.58)
2013	6 (2.83)	8 (7.92)	22 (32.46)	20 (63.65)	31 (107.19)	87	19.73 (15.8,24.34)	10.54 (8.18,13.76)
2014	3 (1.45)	11 (10.82)	24 (34.74)	25 (80.45)	20 (67.55)	83	18.95 (15.09,23.49)	10.19 (7.96,13.27)
			time trend				time trend	
APC (95% CI)	17.89 (-15.97, 67.88)	-6.33 (-22.21, 12.61)	5.96 (-7.77, 21.90)	-1.88 (-13.68, 11.47)	-2.08 (-13.07, 10.32)		0.94 (-5.62,7.95)	
p-value	0.35	0.49	0.41	0.77	0.73		0.79	

Figure 8: *Incidence of Haemorrhagic stroke in Lleida (2010-2014).* Number of stroke cases, Crude Rate (CR) per 100,000 in Lleida from 2010 to 2014 displayed for age groups for male female and both sex population. Global estimation of the cases, CR (95% confidence interval) or ASR (95% confidence interval) for all years also shown for male, female and both sex population. WHO (World Health Organization 2001) standard population used to compute the Age standardized Rate. Annual Percent Change (APC) with 95% confidence interval shown for male , female and both sex population. Poisson models with logarithmic link used to assess the trend and their corresponding statistical significance.

6.1.4 WHO standard population

Age group	w_i
<40	0.6553
40-54	0.1800
55-69	0.1123
$71 - 80$	0.0315
> 80	0.0212

Table 1: Example of the vector (with four age groups) with the *weights* of the standard population proposed by the World Health Organisation in 2001

7 References

[1] J Díaz-Guzmán, JA Egido-Herrero, R Gabriel-Sánchez, G Barberà, and B Fuentes. Bases metodológicas del estudio ibericTus. *Rev neurol*, 47(12):617–623, 2008.

[2] Stephan U Dombrowski, Joan E Mackintosh, Falko F Sniehotta, Vera Araujo-Soares, Helen Rodgers, Richard G Thomson, Madeleine J Murtagh, Gary A Ford, Martin P Eccles, and Martin White. The impact of the uk 'act fast'stroke awareness campaign: content analysis of patients, witness and primary care clinicians' perceptions. *BMC public health*, 13(1):915, 2013.

[3] Jacques Esteve, Ellen Benhamou, and Luc Raymond. Statistical methods in cancer research. volume iv. descriptive epidemiology. *IARC Sci publ*, 128:1–302, 1994.

[4] Timo Hakulinen and Tadeusz Dyba. Precision of incidence predictions based on poisson distributed observations. *Statistics in medicine*, 13(15):1513–1523, 1994.

[5] Peter McCullagh and John A Nelder. *Generalized linear models*, volume 37. CRC press, 1989.

[6] Amanda G Thrift, Dominique A Cadilhac, Tharshanah Thayabaranathan, George Howard, Virginia J Howard, Peter M Rothwell, and Geoffrey A Donnan. Global stroke statistics. *International Journal of Stroke*, 9(1):6–18, 2014.

[7] Joan Valls. *Mortalitat per càncer de mama a Europa l'any 2005 i projeccions per a l'any 2015*. PhD thesis, Facultat de Medicina, UB, 2010.

[8] Ángel Vila Córcoles, Eva Satué Gracia, Olga Ochoa Gondar, Diego Cabanes, Ángel Vila Rovira, Jordi Bladé, Francesc Bobé, Raquel Casanova, Manuel Jariod, Rosaura Reig, et al. Incidencia y letalidad del ictus isquémico en la población mayor de 60 años del área de tarragona, españa (2008-2011). *Rev Neurol*, pages 490–496, 2014.